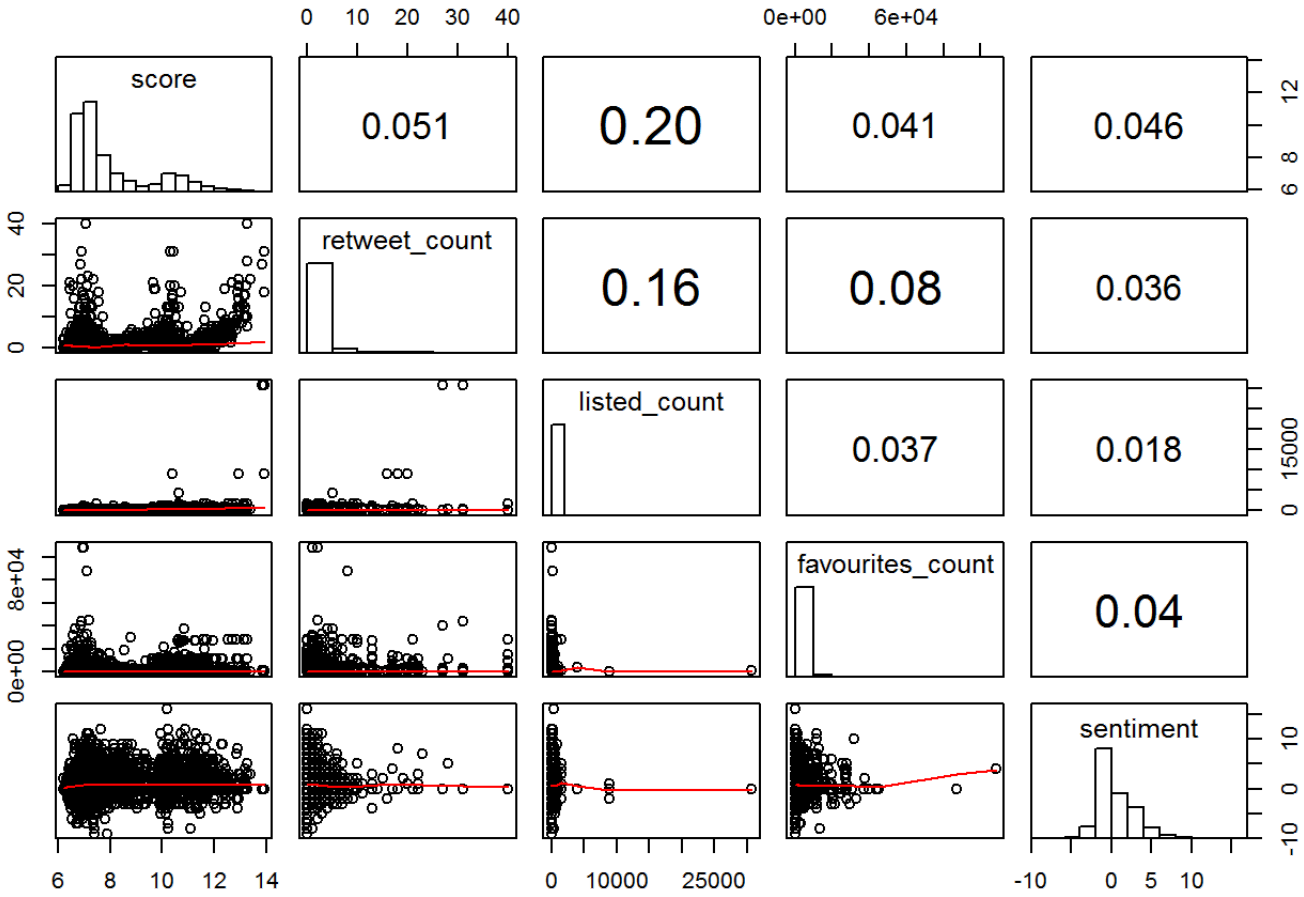
Healthcare Twitter Analytics

Have a look at the relationships among the numerical fields

```
file = 'Tweets_Celiac_sent.csv'
data = read.csv(file)
data=data[c("id","score","retweet_count","listed_count","favourites_count","sentiment")]
first.look(data[c(-1)])
```

```
## ----- str: look at data types and values -------
## 'data.frame':    5675 obs. of  5 variables:
##  $ score           : num  7.73 6.84 6.88 6.86 8.4 ...
##  $ retweet_count   : int  10 0 0 0 2 1 0 0 2 4 ...
##  $ listed_count    : int  142 5 5 5 27 71 71 13 59 223 ...
##  $ favourites_count: int  55 0 0 0 93 0 0 482 62 55 ...
##  $ sentiment       : num  0 0 5 0 3 2 3 3 3 0 ...
## NULL
##
## ----- sample size: 5675
## ----- sample size of complete cases: 5675
## ----- difference: 0
##
## ----- names: variable names
## [1] "score"            "retweet_count"    "listed_count"
## [4] "favourites_count" "sentiment"
##
## ----- some: Randomly select a few elements
##       score retweet_count listed_count favourites_count sentiment
## 194   6.841             5            2             1206         0
## 1140  8.009             1           29               43         4
## 2378  6.840             2           28              638         3
## 3030  7.319             0            1               19         0
## 3086  6.713             1           59               62         1
## 4014  7.737             0            1               97        -2
## 4834  8.277             2           43             1868        -1
## 4842 10.840             1          608              280         3
## 5141  7.231             2            4                2         0
## 5552  7.284             0            7               55         0
##
## ----- summary: statistics for each variable
##      score       retweet_count    listed_count   favourites_count
##  Min.   : 6.22   Min.   : 0.00   Min.   :    0   Min.   :     0
##  1st Qu.: 6.97   1st Qu.: 0.00   1st Qu.:    4   1st Qu.:    36
```

```
##  Median : 7.33   Median : 0.00   Median :   16   Median :   228
##  Mean   : 8.02   Mean   : 1.78   Mean   :   80   Mean   :  1377
##  3rd Qu.: 8.50   3rd Qu.: 2.00   3rd Qu.:   59   3rd Qu.:   998
##  Max.   :13.91   Max.   :40.00   Max.   :30761   Max.   :108340
##    sentiment
##  Min.   :-9.00
##  1st Qu.: 0.00
##  Median : 0.00
##  Mean   : 1.21
##  3rd Qu.: 2.00
##  Max.   :16.00
##
## ----- grfsummary: statistics for each quantitative variable
##            score retweet_count listed_count favourites_count sentiment
## n        5675.00       5675.00      5675.00          5675.00   5675.00
## nNA         0.00          0.00         0.00             0.00      0.00
## min         6.22          0.00         0.00             0.00     -9.00
## mean        8.02          1.78        79.68          1376.79      1.21
## median      7.33          0.00        16.00           228.00      0.00
## stdev       1.55          4.09       627.11          5633.75      2.34
## skew        1.33          4.65        42.81            13.79      0.86
## npskew      0.44          0.43         0.10             0.20      0.52
## kurtosis    0.61         27.33      2036.84           240.79      2.05
## max        13.92         40.00     30761.00        108340.00     16.00
##
##
##  ------- possible factor non-independence ------
## factor variable names:
## List of possible factor pairs with interactions:
```


score
retweet_count
listed_count
favourites_count
sentiment
0.051
0.20
0.041
0.046
0.16
0.08
0.036
0.037
0.018
0.04
0 10 20 30 40
0e+00 6e+04
6 8 12
0 20 40
0 15000
0e+00 8e+04
-10 0 10
6 8 10 12 14
0 10000 25000
-10 0 5 10